Fig 2c: Performance Gap Between Agents by Model